An Artificial Recurrent Neural Network model for Weather Forecasting

[1]M. Swapna, [2]N. Sudhakar

[1]Assistant Professor, [2]Professor
CSE Department & GVPCEW, VSKP CSE Department & BEC, BAPATLA

***Abstract*— Weather forecasting is an application of science and technology to predict the state of the atmosphere for a given location. The available NWP (numerical weather prediction) models do not generally provide forecasts with the accuracy at resolution appropriate for this task. This paper describes the precipitation of weather forecasting based on ANN. Weather forecasting is made by collecting quantitative data about the current atmosphere at a given place and using scientific understanding of atmospheric process to show the changes in the atmosphere. Artificial Neural network is the mimic of human ability which adapts the changing circumstances. It is a complex network with highly interconnected processing elements called Neurons. The Gradient Descent algorithm has been used to predict the visibility of Coastal Andhra Pradesh based on different parameters. The data of these parameters is normalized initially and then trained. Further the performance of network is compared and the weather parameter is predicted.**

Keywords*— *Recurrent Artificial Neural Networks, Back propagation Algorithm, Gradient Descent strategy, Precipitation, Prediction.

I. INTRODUCTION

Meteorological services often perform complicated and challenging tasks all over the world. The forecasting methodologies are separated into two branches in terms of numerical modeling and scientific processing by researchers in this field. Success is rarely visible using the numerical method though research is conducted for a long time. The statistical method is used as an alternative in which parameters are treated as stochastic is widely used for prediction. Numerical weather prediction (NWP) models have significantly increased the accuracy of forecasts and similar improvements in the accuracy of precipitation forecasts have not been attained because of the physical complexity of precipitation processes.

Gradient Descent is the first order algorithm which is used for local minimum of a function. It takes as much iteration it requires to minimize for the accuracy. For supervised learning, the networks are with vectors of parameters in them. Multiple linear regression is used to the parameters, probability of precipitation, and other variables in order to interest at specific locations and these derived relationships are in turn used to predict the variables from real-time NWP model output.

Artificial neural networks have expanded and have seen an explosion of interest over the last few years. This is successfully applied to a wide range of problem domains which include areas in diverse as finance, medicine, engineering, etc.

II. ARTIFICIAL NEURAL NETWORK (ANN)

The Artificial Neural Network is an interconnected network of neurons to pass information using electric signals. Its architecture is made up of an input, output and one or more number of hidden layers. The nodes are internally connected from the input layer, hidden layers and the output layer with some weight associated with it. Raw information which is fed into the network as input never changes its value[10]. The input data is duplicated and sent to the hidden layers where the weight is modified and sent to the output layer after the weight modification in hidden layer. The network starts evolving, neurons continuously evaluate their output by looking at their inputs, calculating the weighted sum and comparing to a threshold to decide if they should fire. The dependence of output values on input values is quite complex and includes all synaptic weights and thresholds[10]. Recurrent neural network is used as it uses internal memory to process the arbitrary inputs.

III. PREDICTION USING ANN

Recurrent neural networks employ feedback connections and have the potential to represent certain computational structures[1].This is used to solve machine learning problems. Gradient descent starts with initial values and iteratively moves towards a set of parameter value that minimize the function. This iterative minimization is achieved using calculus. At the staring the network is trained and then tested. The network is produced with initial weights which are the inputs and outputs. To present the magnitudes of both inputs and outputs in the same magnitude, they are normalized to an order of scale 0 to 1. Linear Perceptron in neural network[4] is similar to the behaviour of network of activation functions. Back Propagation algorithm which is used in conjunction with gradient descent method is the method used in prediction of weather forecasting. This method requires computation of error function at each iteration and the continuity and differentiability is to be guaranteed. So, we need to use an activation function other than the step function as the composite function and the interconnected perceptrons are discontinued. Among different activation functions, the most commonly used are the Sigmoid, Threshold, Hyperbolic tangent functions. The hyperbolic and the sigmoid functions are the same kind of differentiable functions with the only difference as the outputs of hyperbolic tangents range from -1 to 1 and the outputs range from 0 to 1. A mathematical function which refers to a special case of logistic function[4,5] is the Sigmoid function. The 'S' shaped graph of sigmoid function is shown in the given figure.

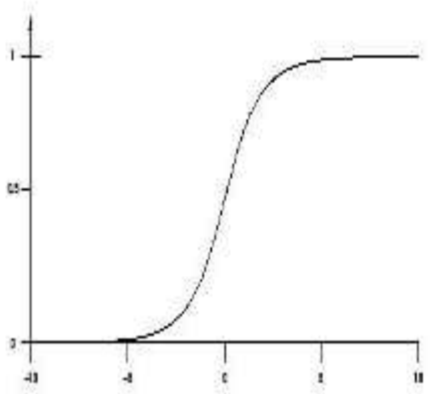

Fig1: Graph of Sigmoid function

The Sigmoid function equation which is real valued and differentiable is given below.

$$S(t) = \frac{1}{1 + e^{-t}}$$

'S' is the sigmoid function, 'e' is the natural logarithm and 't' is the steepness of curve.

IV. BACK PROPAGATION ALGORITHM

Back Propagation learning strategy is the standard recurrent network that performs Gradient Descent on a complete unfolded network. The information processing system that produces an output value with on one or more inputs. On repeated presenting theknown input and outputs to the network, it reduces the mean squared error with each iteration of the training. The back propagation is an expression in the network[6][10] which is the partial derivative of the cost function 'C' with respect to any weight 'W' or bias ' b' in the network.

The minimum error function[7][10] in weight space using the method of gradient descent method is the main aim of back propagation. The solution of the learning problem is the combination of weights which minimizes the error function. In order to compute the necessary corrections after choosing the weights in the network, the Back Propagation algorithm is used. The algorithm is divided into four steps. They are:

1. Feed-forward computation
2. Backpropagation to the output layer
3. Backpropagation to the hidden layer
4. Weight update

Feed-Forward Computation:

The forward propagation of the inputs to the network is processes to produce the output activations. These output activations are compared with the desired outputs and the error is propagated backward to the previous layers.

Backpropagation to the output layers:

The training pattern target is used to propagate the output activations through the network. This generates the deltas of the output neurons where delta is the difference between the targeted and the actual output.

Backpropagation to the hidden layers:

The same training pattern target is used to propagate the output activations from the output layer. Here the delta value is calculated of all the hidden layer neurons.

Weights update:

For updating the weights, we follow two steps.

a) The output delta and the input activations are multiplied to get the gradient of the weight.
b) The ratio(percentage) is subtracted from the weight of the gradient

The ratio (percentage) which is called the learning rate influences the speed and quality of learning. From this ratio we can conclude that greater the ratio greater the faster the neuron gets trained and lower the ratio, the more accurate is the training. Finally, we repeat the four steps until performance is satisfactory in the network.

V. GRADIENT DESCENT STRATEGY

One of the most popular algorithms to perform optimization in neural networks. This minimizes the objective function which is written as the sum of differentiable functions. The gradient descent method which is also known as the steepest descent[8][10] is the iterative method. Implementation of large scale optimization problems is the popular characteristic of gradient descent strategy as it can handle "black box" functions. Starting with an initial set of parameters and moving iteratively to minimize the function. By taking the steps in negative direction of the function gradient the iterative minimization is achieved. This mathematical explanation is sometimes difficult. So, this is translated into practical settings. One of the canonical example is the Linear Regression to explain Gradient descent. This Gradient descent strategy is used to solve machine learning problems like linear regression.

Linear Regression:

The main goal is to form a line using a set of points. The graph below shows the linear regression by gradient descent.

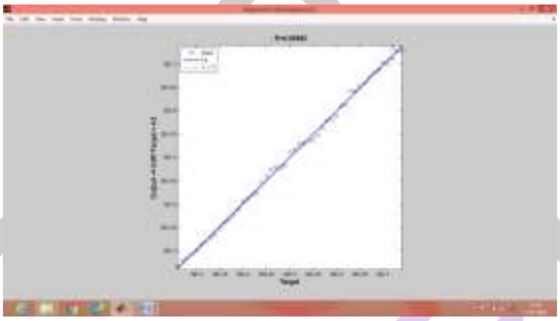

Fig2: Linear Regression

If we want to model the given points, we will use the standard line equation $y = mx + c$ where m is the line slope and c is the y-intercept. The best slope value and y-intercept are to be found to form the best line. For solving this type of problems, there is a standard approach that is to define the error function which is also known as the cost function. The input pair (m ,c) is given which shows the error value and the see how well the line fits. We compute through each (x ,y) point in our data set. The sum of the squared distances is calculated to make our error function differentiable.
The error function looks like:

$$Error(m,c) = 1/N \sum_{i=1}^{N} (yi - (mxi + c))2$$

where m and c are the parameters which are called the slope and intercept.

VI. MODEL REPRESENTATION AND EXPERIMENTAL RESULTS

Four different parameters are considered and the data for the same is collected from to 2015. The four parameters include the wind, pressure, the minimum and maximum temperatures. The data is normalized and is taken for training the network. At the same time different data is taken for testing. The results obtained are comparatively similar. The model has one input layer, many hidden layers and one output layer with the learning rate take as 1. The testing data is presented to the network and is compared with the actual data. The Mean Squared Error (MSE) is calculated. If the output is approximately equal to the actual data, then it is the predicted value. The model of the network is shown below.

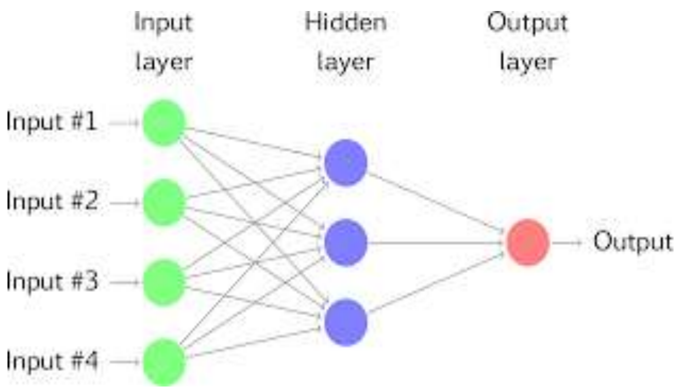

Fig3: Model of the network

The data of the parameters which are given as input is shown in the following table:

Table1: INPUT DATA FOR THE GIVEN FOUR PARAMETERS

DATE	MIN TEMP	MAX TEMP	WIND	PRESSURE
02/01/13	30	21.2	2.8	1012.9
03/01/13	28.8	20.6	4.6	1013.2
04/01/13	30	21.4	4.2	1012.9
05/01/13	30	21.5	1.9	1012.3
06/01/13	29	22.4	6.9	1011.5
07/01/13	29	21.8	5.6	1012.4
08/01/13	29.2	22	4.2	1013.2
09/01/13	29.7	20.7	9.7	1013.8
10/01/13	29.7	21.4	6	1014.4
11/01/13	28.3	19.6	5.1	1014
12/01/13	28.3	19.4	6.3	1014.7
13/01/13	28.2	19	3.9	1014.5
14/01/13	28.8	17.8	6.5	1013.4
15/01/13	28.8	18.9	3.9	1013.7
16/01/13	28.6	18.8	2.5	1015.4
17/01/13	28.6	19.6	2.1	1015.9
18/01/13	30.2	19	8.1	1016.4
19/01/13	30.2	19.9	6.3	1016.2
20/02/13	29.9	19.4	4.9	1015.4
21/01/13	29.1	19.2	4.2	1015.4
22/02/13	29.1	19.4	6.7	1017
23/01/13	29.3	19.6	6	1016.8
24/01/13	29.3	21	6.5	1016.8
25/01/13	28.5	19.7	8.8	1016.2
26/01/13	29.4	20.4	7.6	1016.4
27/01/13	29.4	20.4	3.9	1016.4
28/01/13	28.4	20.4	2.8	1016
29/01/13	28.4	20	3.7	1015
30/01/13	28.4	20.5	4.9	1014.6

The resultant performance of the graph looks as follows.

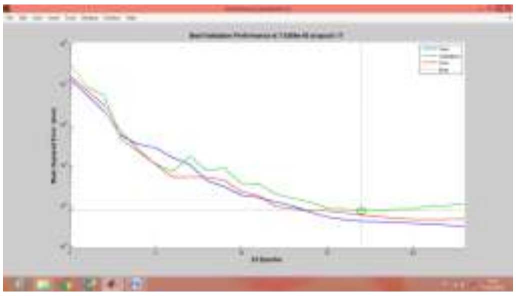

Fig4: Performance graph

VII. CONCLUSIONS AND FUTURE ENHANCEMENTS

Weather predictions overcome the most significant socio-economic impact over many parts of the world. With the recent advancements in neural network methodology for modeling, along with a wide range of applications are motivated to find the applications of ANN in weather prediction. In this paper, the developed ANN model with the Backpropagation and the gradient descent strategies are applied to predict the visibility. This uses few other parameters of weather data of coastal Andhra Pradesh. The results predicted have shown the efficiency of ANN. Finally, it can be inferred that ANN can yield more accurate results provided when the selection strategies, training paradigms and the network input, outputs are determined properly. In future, i would like to observe different other weather parameters to predict the weather with the same algorithm. This paper gives the study of recurrent neural networks getting trained by Backpropagation in conjunction with gradient descent strategy which is for the efficient predictions of Humidity.

REFERENCES

[1] "Noisy time series prediction using recurrent neural network and grammatical inference" by C. Lee Gilesand Steve Lawrence.
[2] Mordecai Avriel(2003) Nonlinear programming: Analysis and methods.
[3] Comparison of neural network configurations in the prediction of monthly rainfall over coastal Andhra Pradesh- base paper.
[4] Haykin, Simon(1998), Neural networks: A comprehensive foundation (2 ed)
[5] Han Jun, Morag, Claudio(1995), The influence of sigmoid function parameters on the speed of back propagation learning, In Mira, Jose, Sandovol, Francisco.
[6] Neural Networks and deep learning.com
[7] Berlin.de.rojas neural network
[8] Homes.soic.indiana.edu/classes/spring2012/csci/b553-hauserk/gradient descent.
[9] Hopfield, J. [1990a], "Artificial Neural Networks are Coming," IEEE Expert An Interview by W. Myers, pp. 3-6, April.
[10] Draye, J.P., Pavisic, D., Cheron, G. and Libert, G. [1997], "An Inhibitory Weight Initialization Improves the Speed and Quality of Recurrent Neural Network Learning," Neurocomputations, Vol. 16, pp. 207-224.Dreyfus, S.E. [1990], "Artificial Neural Networks, Back Propagation and the Kelley-Bryson Gradient Procedure," J. Guidance, Control Dynamics, Vol. 13, No. 5, pp. 926-928.
[11] Lee, B.W., Liu, J. and Dai, H. [1998], "Forecasting From Low Quality Data with Applications in Weather Forecasting," Inter. J. Comput. Inform., Vol. 22, No. 3, pp. 351-358.